PHIL 7001: Fundamentals of AI, Data, and Algorithms

Week 5B
Bayesian Statistics

Boris Babic,
HKU 100 Associate Professor of Data Science, Law and Philosophy

Today

Learning goals

- Introduction to Bayesian Inference
- Understanding Prior, Likelihood, and Posterior
- Challenges in Bayesian Calculations
- Beta-Binomial Model and Its Intuitive Interpretation
- Making Inferences from Bayesian Models
- Practical Shortcuts in Bayesian Inference

Review from last class

Last class

Reviews of last class

- Population, Sample, Point Estimates
- Hypothesis tests
- Significance level
- P-value

The Bayesian Approach

Frequentist vs. Bayesian Statistics

- Frequentist approach:
    - Assume parameters of interest are fixed.
    - Often look at long-run.
- Bayesian approach:
    - Assume parameters follow some distribution (prior).
    - Update based on data and prior knowledge.

Introduction to Bayesian Inference

- **Prior**: Initial belief about a parameter, based on existing knowledge.
- **Likelihood**: The probability of observed data under various parameter values.
- **Posterior**: Updated belief after observing data. Calculated using Bayes' Rule:

$$P(\theta|data) = \frac{P(data|\theta) \cdot P(\theta)}{P(data)}$$

- Where θ is some (population) parameter of interest.
- **Inference**: Making decisions or predictions based on the posterior.

Bayesian Framework

- Posterior distribution of θ (conditional distribution of θ given $\boldsymbol{X}$):

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)}{m(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{m(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta)\pi(\theta) \qquad \text{(Bayes' Rule)}$$

- This looks exactly like the Bayes' Theorem that we have learned, but now, probabilities of events have been replaced with general distributions.
- This is hard!! (do not be scared!!) So we will use shortcuts!!

The Beta Binomial Strategy

Problem

A Curious Coin

You have come across a curious coin. It seems (you suspect) bent in a way that biases it toward landing on heads. You will give this coin to your trusty RA, and ask them to perform an experiment (i.e., toss it repeatedly) in order to help you decide whether the coin is biased.

Some Data

A Curious Coin

Your RA reports having tossed the coin 12 times, with the following results:

$$H, T, H, H, H, H, H, T, H, H, H, T$$

$$9H, 3T$$

Inference

- This is an instance of a problem about inference for a proportion, θ.
- Some questions you might want to ask include:

  What is a good point estimate of θ?

  How confident am I that θ is in some range $[a, b]$?

  Can I reject the hypothesis that $\theta = 0.5$?
- Before you can answer these questions you have to ask:

  What is the prior distribution of θ?

  What is the distribution of the data – i.e. of the coin tosses?

  What is the posterior distribution for θ?

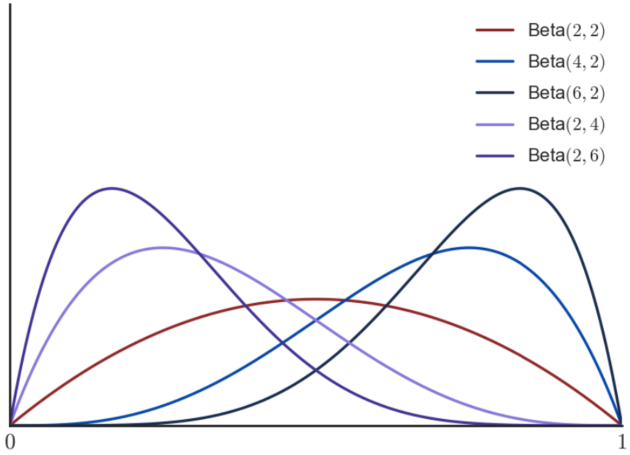
Distribution of the data

- This is the easiest for us to answer.
- The data comes from tossing a coin.
- So the distribution of the data is binomial.
- Hence: $f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$.
- A trick: For Bayesian calculation, we can ignore terms that are not functions of θ, because we are ultimately interested in θ.
- Therefore: $f(x) \propto \theta^x(1-\theta)^{n-x}$.
- The infinity-like symbol means "proportional to". It means that the two terms on the left and the right are equal, up to multiplication by some constant value that is not a function of θ.

Prior for θ

- This is an art. So you will have to trust me!
- We want to choose a prior that is flexible, convenient, and easy to make calculations with.
- Also, remember that θ is a proportion. So we are looking for a probability distribution over a proportion. A proportion can be only between 0 and 1, so the best prior would be a prior that can only take values in 0 and 1.
- Let us choose the following: $\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.
- This is called a beta prior. A beta distribution is a distribution for a quantity that takes values between 0 and 1, and that is exactly what we need here.
- The quantities α and β are the parameters of the prior distribution.
- Notice that I picked a prior, which mimics (looks very similar to) the binomial data distribution.

Beta prior



Posterior for θ

- We said the prior is $\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.
- We said the data distribution is $f(x) \propto \theta^{x}(1-\theta)^{n-x}$.
- The posterior is given by: $\pi(\theta|x) \propto f(x) \times \pi(\theta)$.
- If we multiply the above prior, by the above data distribution, we obtain: $\pi(\theta|x) \propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$
- Meaning: the parameters α and β correspond to the number of heads, and tails, that you ASSUME you observed, before seeing the data. They are like fake or imaginary tosses! This is YOUR prior.
- Example: Start with a prior distribution which imagines 1 Heads, and 1 Tails.
- Another example: Start with a prior distribution which imagines 75 Heads, and 90 Tails.
- Which approach is better? Why?

Bayesian Approach to the Coin Problem

- Let $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$. Let $\beta^* = n - \sum_{i=1}^{n} x_i + \beta$. Then our posterior distribution for θ is

$$\pi(\theta|\boldsymbol{x}) \propto \theta^{\alpha^*-1}(1-\theta)^{\beta^*-1}$$

- In other words, the posterior distribution is still of the beta form, except that our new α corresponds to the initial α plus the number of successes/heads and the new β corresponds to the initial β plus the number of failures/tails.
- So again, to repeat: this lends itself to a natural interpretation: The initial α value corresponds to the number of pseudo tosses that came up heads, whereas the initial β value corresponds to the number of pseudo tosses that came up tails.
- Bayesian updating is easily accomplished by adding the pseudo heads to the observed heads and pseudo tails to observed tails.

Conjugate Families

- What we have seen so far is that the beta distribution is conjugate to the binomial distribution. This is sometimes called the "beta-binomial" family.
- What it means is that if we start with a beta prior, and update on binomial data, we end with a beta posterior!
- This makes it very easy to do Bayesian inference with minimial calculations. Let's see how!

Some properties of the Beta distribution

- Suppose X follows a $\mathrm{Binomial}(n, \theta)$ distribution.
- In order to make predictions about θ or about new observations in the future, we will use the posterior mean. That's it!
- Example 1: What is the prevalence of the flu in Hong Kong right now? This is a question about the rate itself.
- Example 2: What is the probability that Cindy, a student at HKU, has the flu right now? This is a prediction about a particular individual.
- For both Example 1 and Example 2, we use the posterior mean of θ.
- Prior mean: $\alpha/(\alpha + \beta)$
- Posterior mean: $(\alpha + x)/(\alpha + \beta + n)$
- Example: we start with a prior where $\alpha = 1, \beta = 1$. We observe 10 people, 7 have the flu. 3 do not.
- Posterior is also beta with new α given by 7 + 1 = 8. And new β given by $3 + 1 = 4$.
- Posterior mean is: 8/(8 + 14) = 8/12. That's it!
- Compare: if we were non Bayesians, the sample mean that we use to make an estimate would be 7/10.

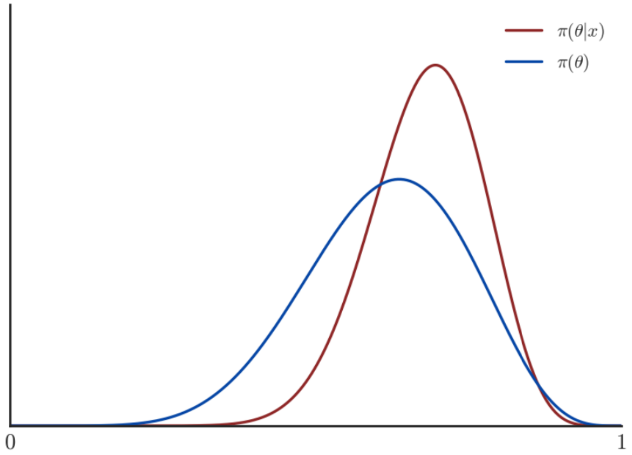
The Beta-Binomial Model - Intuition

- Picture this as the coin-flipping scenario.
- **Beta distribution**: Our initial belief about the fairness of the coin.
- **Binomial data**: The observed coin flips.
- **Posterior**: Updated belief about the fairness of the coin.
- Beta parameters can be thought of as "prior number of successes and failures".
- Example: Beta(9, 3) might represent a prior belief based on 9 heads and 3 tails.

Bayesian Approach to the Coin Problem

- Let's return to our coin problem.
- We said the coin looked bent in a way that made it biased toward heads (i.e., $\theta > 0.5$).
- What are reasonable values for α and β?
- Suppose $\alpha = 8$ and $\beta = 5$.
- The posterior distribution is beta with $\alpha = 9 + 8 = 17$ and $\beta = 3 + 5 = 8$.

Posterior Distribution



Bayesian Hypothesis Tests

Bayesian Hypothesis Tests

- Recall that what we really wanted to know was a simple question: is the coin biased toward heads?
- Now we can answer it directly:

$$\begin{aligned}\Pr(\theta > 0.5) &= \int_{0.5}^{1} \pi(\theta|\boldsymbol{x})d\theta \\ &= 1 - CDF(\theta|\boldsymbol{x})|_{\theta=0.5} \\ &= 1 - 0.03 \\ &= 0.97\end{aligned}$$

- R code: `1 - pbeta(0.5, 17, 8)`
- We are 97% confident that the coin is biased toward heads.
- We now have an answer to a one-sided hypothesis test:

$$H_0 : \theta \leq 0.5 \qquad\qquad H_1 : \theta > 0.5$$

- But instead of accepting/rejecting the null hypothesis, we make probabilistic statements about the research hypothesis from the posterior distribution.

Credible Intervals

Credible Intervals

- A 95% credible interval for θ is,

$$\Pr(a < \theta < b) = \int_a^b \pi(\theta|\boldsymbol{x})d\theta = 0.95$$

- In our case, $a = 0.49$ and $b = 0.84$.
- R code: `qbeta(c(0.025,0.975),17,8)`
- We can also compute the probability that θ is in any desired region of the posterior distribution. This gives us a probabilistic statement about a small region around a point null hypothesis. For example:

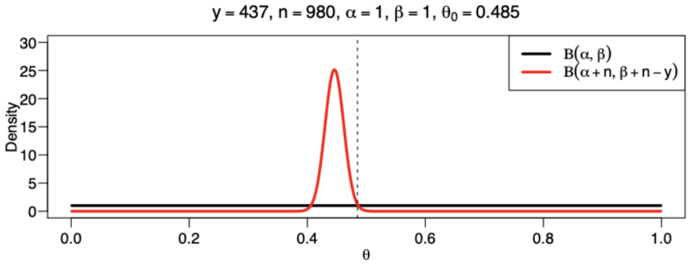
$$\begin{aligned}\Pr(0.4 < \theta < 0.6) &= \int_{0.4}^{0.6} \pi(\theta|\boldsymbol{x})d\theta \\ &= CDF(\theta|\boldsymbol{x})|_{\theta=0.6} - CDF(\theta|\boldsymbol{x})|_{\theta=0.4} \\ &= 0.19\end{aligned}$$

- R code: `pbeta(0.6, 17, 8) - pbeta(0.4, 17, 8)`.
- We are about 20% confident that θ is between 0.4 and 0.6.
- Compare this to the confidence interval from Class 1A. This is now a probabilistic statement about θ, treated as a random quantity. And not a probabilistic statement about X and the proportion of cases in which it will cover θ if sampled repeatedly!

Example: Placenta Previa

- A study was conducted in Germany of 980 births from women with placenta previa. Out of the 980 births, $y = 437$ were baby girls.
- Note: $X \sim$ Bernoulli(θ) and $Y = \sum X_i \sim$ Binomial(n, θ)
- The established proportion of female births in the general population is 0.485.
- The scientific question of interest is whether the proportion of female births in this subpopulation is less than that in the general population.
- Let θ denote the proportion of female births.
- Assume $\theta \sim$ Beta$(1, 1) =$ Uniform$(0, 1)$.
- Find $\theta|y$
- Find $\mathbb{E}(\theta|y)$
- Find 95% credible interval of $\theta|y$

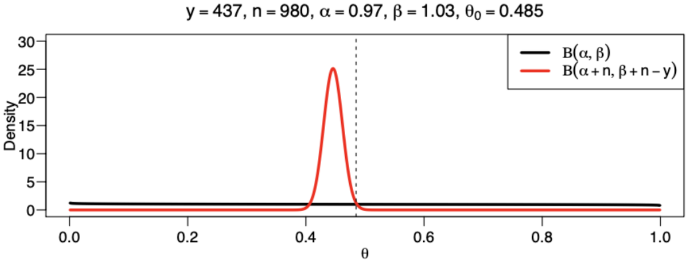
Example: Placenta Previa



$E(\theta)$	$\alpha+\beta$	$E(\theta \mid y)$	95% Credible Interval
0.5	2	0.446	(0.415, 0.477)
0.485	2	0.446	(0.415, 0.477)
0.485	20	0.447	(0.416, 0.478)
0.485	200	0.453	(0.424, 0.481)

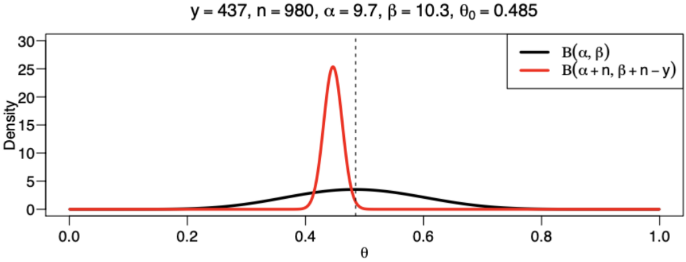
Note that $y/n = 0.445$.

Example: Placenta Previa



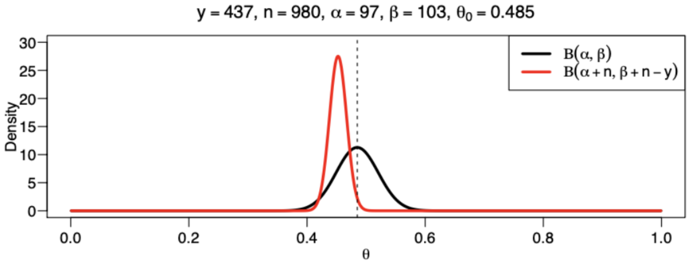
$E(\theta)$	$\alpha + \beta$	$E(\theta \mid y)$	95% Credible Interval
0.5	2	0.446	(0.415, 0.477)
0.485	2	0.446	(0.415, 0.477)
0.485	20	0.447	(0.416, 0.478)
0.485	200	0.453	(0.424, 0.481)

Example: Placenta Previa



$E(\theta)$	$\alpha + \beta$	$E(\theta \mid y)$	95% Credible Interval
0.5	2	0.446	(0.415, 0.477)
0.485	2	0.446	(0.415, 0.477)
0.485	20	0.447	(0.416, 0.478)
0.485	200	0.453	(0.424, 0.481)

Example: Placenta Previa



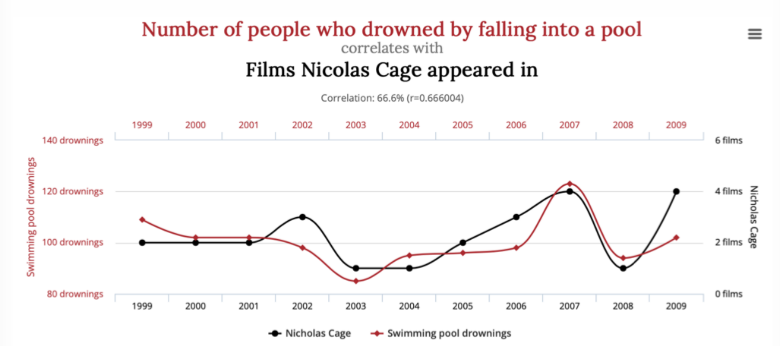
$E(\theta)$	$\alpha + \beta$	$E(\theta \mid y)$	95% Credible Interval
0.5	2	0.446	(0.415, 0.477)
0.485	2	0.446	(0.415, 0.477)
0.485	20	0.447	(0.416, 0.478)
0.485	200	0.453	(0.424, 0.481)

Today

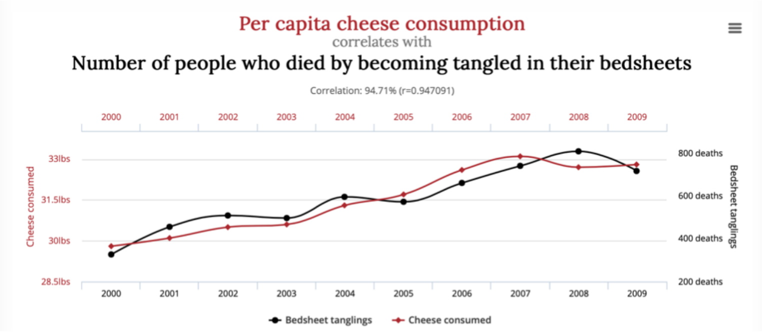
- Introduction to Bayesian Inference
- Understanding Prior, Likelihood, and Posterior
- Challenges in Bayesian Calculations
- Beta-Binomial Model and Its Intuitive Interpretation
- Making Inferences from Bayesian Models
- Practical Shortcuts in Bayesian Inference

Next week

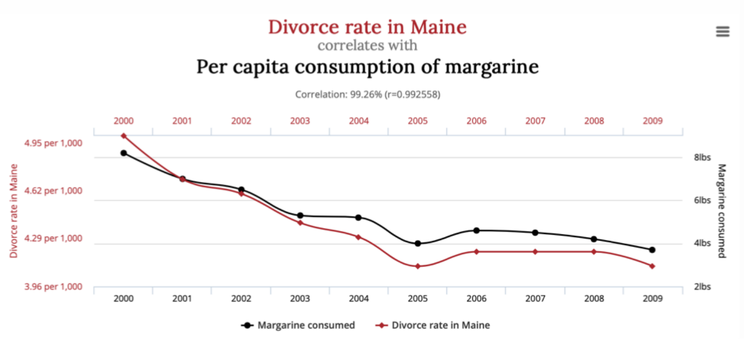
- Linear regression
- Intuition for simple linear regression models
- Using R to build linear regression models
- Evaluating output from the models
- Checking whether a model is good


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
Correlation: 66.6% (r=0.666004)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings

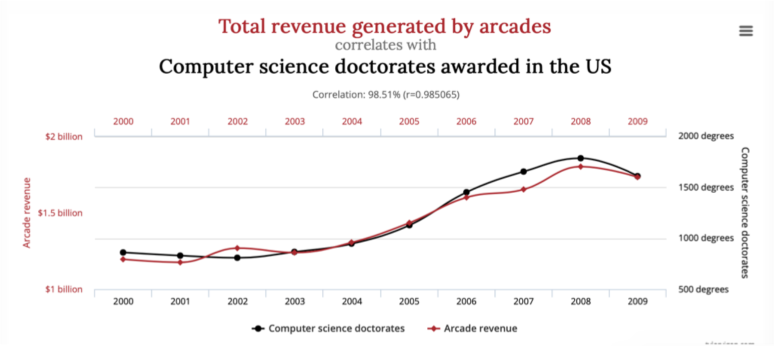
Spurious Correlations



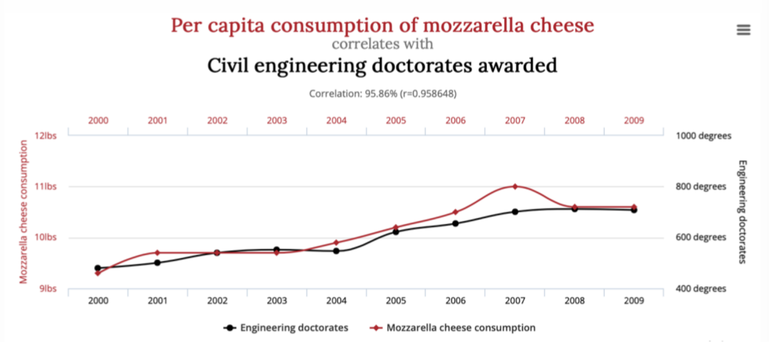
Spurious Correlations



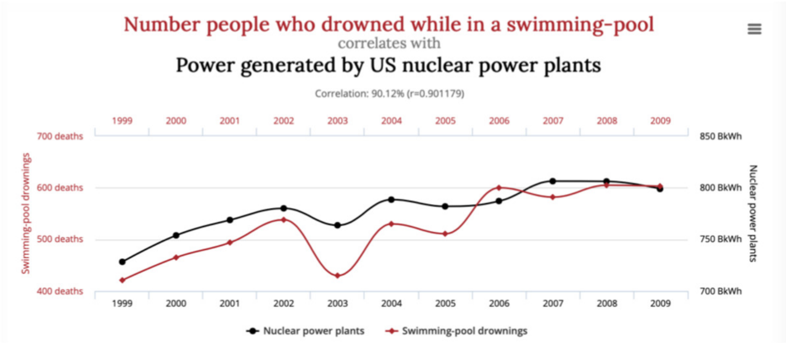
Spurious Correlations



Spurious Correlations



Spurious Correlations



Spurious Correlations